Training in Biomedical Informatics at Columbia University

Columbia University

PI: HRIPCSAK, GEORGE M. Grant Number: 5T15 LM007079-24S1

We propose a new track in our training program focused on Big Data Science applied to biomedicine and health at Columbia University. This will involve creation of a new course in Big Data Science as well as a new specialized track in Biomedicine and Health Data Science for pre-doctoral degree students. The proposed new track will supplement the existing training opportunities for our students. With the escalating cos of healthcare and problems in the delivery of quality care, there is a vital need to develop high-throughput methods specific to healthcare as they have the potential to revolutionize medicine and biomedical research. With the explosion of biomedical knowledge and health-related data coming from the literature, the Internet, and the Electronic Health Record there is an opportunity to develop new methods that incorporate massive amounts of data and knowledge to derive new biomedical knowledge, and then to use the knowledge to improve human health. For approximately the last twenty-five years, new graduate programs at a limited number of Universities have been created to train researchers and professionals in Biomedical Informatics programs, but in these programs only a small component involve training in use of big data. Similarly, in recent years, graduate level programs have been created to train researchers and professionals in techniques based on the use of large amounts of complex heterogeneous data. However, the focus of these programs is not on medicine and health care. We see a need at Columbia University to create a curriculum that leverages courses related to Data Science and research from multiple disciplines and apply the knowledge to biomedical and health related applications. Our curriculum will include a course in reproducibility given in our department that will involve projects in biomedicine and health as well as using existing courses in Computer Science, Data Science, Statistics, and Biostatistics so that students learn the relevant methods. The courses will be supplemented with joint mentorship in research opportunities geared specifically toward big data as applied to biomedicine and health. This proposal is an excellent opportunity for Columbia University. It draws on the strengths of our graduate program in Biomedical Informatics Columbia, a new graduate program in Data Science initiated in 2013, as well as the participation of well-known faculty in multiple fields. Our existing program is well-established, attracts strong applicants, and has an impressive track record that includes years of experience in training next-generation researchers, many of whom have become leaders in academic, industry, government, and health care institutions. PUBLIC HEALTH RELEVANCE: The development of a new track in Big Data Science as applied to biomedicine and health will train a new generation of scientists and professionals in using big data to advance medicine, health, and healthcare. Currently, there is a vital need to develop high-throughput methods specific to healthcare as they have the potential to revolutionize medicine and biomedical research by using the massive amounts of biomedical knowledge and health-related data coming from the literature, the Internet, and the Electronic Health Record to derive new biomedical knowledge, and then to use the knowledge to improve human health.